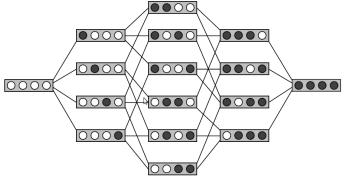
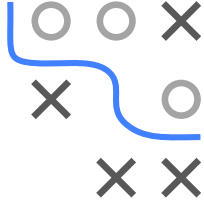
Supervised Learning

Practical Tips for Feature Selection

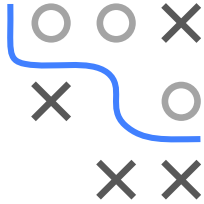


Learning goals

- Add learning goals

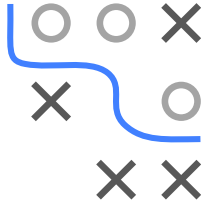


SOLVING A FEATURE SELECTION PROBLEM (TAKEN FROM GUYON (2003))



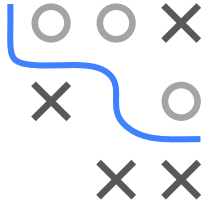
1. **Do you have domain knowledge?**
   If yes, construct a better set of “ad hoc” features.
2. **Are your features commensurate?**
   If no, consider normalizing them.
3. **Do you suspect interdependence of features?**
   If yes, expand your feature set by constructing conjunctive features or products of features, as many as your computer resources allow you to.
4. **Do you need to prune the input features (e.g. for cost, speed or data-understanding reasons)**?
   If no, construct disjunctive features or weighted sums of features (e.g. by clustering or matrix factorization).

SOLVING A FEATURE SELECTION PROBLEM (TAKEN FROM GUYON (2003))



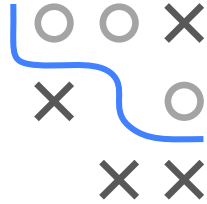
5. **Do you need to assess features individually (e.g. to understand their influence on the system, or because their number is so large that you need to do a first filtering)**?
   If yes, use a variable-ranking method. Otherwise, do it anyway to get baseline results.
6. **Do you need a predictor?**
   If no, stop.
7. **Do you suspect your data is "dirty" (has a few meaningless input patterns and/or noisy outputs or wrong class labels)?**
   If yes, detect the outlier examples using the top-ranking features obtained in step 5 as representation; check and/or discard them.

SOLVING A FEATURE SELECTION PROBLEM (TAKEN FROM GUYON (2003))



8. **Do you know what to try first?**
   If no, use a linear predictor. Use a forward selection method with the "probe" method as a stopping criterion or use the $L0$ norm embedded method. For comparison, following the ranking of step 5, construct a sequence of predictors from the same family, using increasing subsets of features. Can you match or improve performance with a smaller subset? If yes, try a nonlinear predictor with that subset.
9. **Do you have new ideas, time, computational resources, and enough examples?**
   If yes, compare several feature selection methods, including your new idea, correlation coefficients, backward selection and embedded methods. Use linear and nonlinear predictors. Select the best approach via model selection.

SOLVING A FEATURE SELECTION PROBLEM (TAKEN FROM GUYON (2003))



10. **Do you want a stable solution (to improve performance and/or understanding)?**
    If yes, sub-sample your data and redo your analysis for several bootstraps.

OPEN POINTS / PROBLEMS

- In general, it is difficult to give suggestions on when to use which feature selection method.
- Most of the time, it is reasonable to start with a simple, fast method. If this yields unsatisfactory results, one can gradually move to more expensive methods.
- Not every introduced method can be generalized to multi-class problems in an easy fashion.
- Combining the choice of an appropriate classifier and parameter tuning with feature selection is not simple.